BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e141114
doi: 10.3897/biss.8.141114

OPEN ACCESS

Conference Abstract

Advancing Data Standardisation: OBIS Australia's Contribution to Marine Biodiversity Data Publishing in Australia

Sachit Rajbhandari‡, Katherine Tattersall‡, Dave Watts‡

‡ CSIRO, Hobart, Australia

Corresponding author: Sachit Rajbhandari (sachit.rajbhandari@csiro.au)

Received: 06 Nov 2024 | Published: 06 Nov 2024

Citation: Rajbhandari S, Tattersall K, Watts D (2024) Advancing Data Standardisation: OBIS Australia's Contribution to Marine Biodiversity Data Publishing in Australia. Biodiversity Information Science and Standards 8: e141114. https://doi.org/10.3897/biss.8.141114

Abstract

OBIS Australia (OBIS-AU), a regional node of the Ocean Biodiversity Information System (OBIS), promotes the use of Darwin Core (DwC) (Wieczorek et al. 2012) as a standard for marine biodiversity data publishing across the Australian region. OBIS-AU is hosted by Commonwealth Scientific and Industrial Research Organisation (CSIRO) National Collections and Marine Infrastructure (NCMI) and has published over 36 million DwC records from 480 marine biodiversity datasets, establishing itself as a key contributor to the OBIS network. The OBIS-AU publishing workflow leverages centralised data transformation processes to ensure high-quality, standardised data outputs as illustrated in Fig. 1.

OBIS-AU collates new datasets from the CSIRO Marine National Facility (MNF) and from national marine research partners. In addition, OBIS-AU identifies relevant open access marine biodiversity datasets through literature monitoring and links in scientific journal publications and Australian marine researchers are identified by their Open Researcher and Contributor ID (ORCID). Additional key data sources for OBIS-AU include Australian university repositories and global data portals such as Dryad, Zenodo, Pangaea, and GlobalArchive. A crucial step in the publication process involves liaising with data owners

© Rajbhandari S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

to obtain their consent for publication and allowing them to review their datasets before publication.

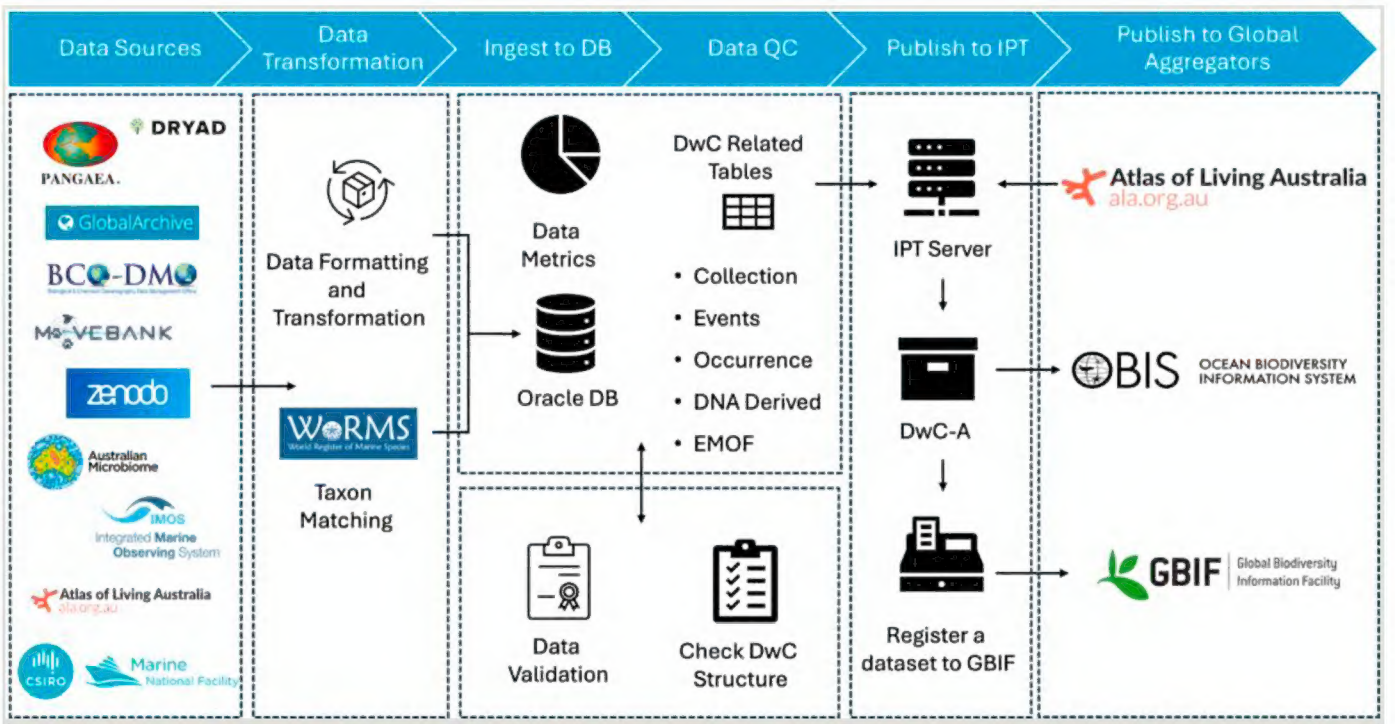

Figure 1.

OBIS-AU Data Publishing Workflow

The OBIS-AU workflow transforms collated datasets into the DwC standard. In conjunction with the Event or Occurrence core, the Extended Measurement or Facts (eMoF) extension is used to store biotic and abiotic measurements or facts related to events or occurrences and linked where possible to formal vocabularies (De Pooter et al. 2017). As OBIS uses the World Register of Marine Species (WoRMS) for its taxonomic backbone, OBIS-AU also identifies the scientific name and retrieves the scientificNameID by matching the taxon name to the WoRMS database.

The OBIS-AU workflow derives metrics from the data and stores them for spatial, temporal, and taxonomic classification accuracy, completeness, and suspect values in the OBIS-AU database to facilitate pre-publication data quality checks. These metrics utilise various data quality criteria established by OBIS and TDWG Task Group 2 Data Quality Tests and Assertions (Belbin et al. 2020). Maps and time series graphs are created for visual inspections of any data outliers. Outliers can be corrected or can be flagged as not-for-publication. The OBIS environmental data (OBIS-ENV-DATA) structures, which include a DwC Event Core along with a DwC Occurrence extension and the ExtendedMeasurementOrFact extension, are checked to ensure that all links between the DwC structures are valid and that there are no orphan records.

OBIS-AU proactively integrates developing data standards into data workflows. In internal data systems only, OBIS-AU associates datasets with applicable Global Ocean Observing System (GOOS) Essential Ocean Variables (EOV). To date, OBIS-AU has linked 227 datasets to *EOV: Fish abundance and distribution* and 21 million occurrence records to *EOV: Microbe biomass and diversity*. Also in leading adoption of developing data standards, OBIS-AU has published 23 eDNA datasets comprising 21 million records

using a DNA-derived data extension in conjunction with the Occurrence core, in accordance with publishing guidelines from the Global Biodiversity Information Facility (GBIF) (Abarenkov et al. 2023). However, the current data model and infrastructure do not support the use of DNA-derived data extensions with Event core. Consequently, event-level measurements must at present be linked to each occurrence record resulting in data redundancy. To avoid data redundancy, OBIS-AU currently publishes event-level measurement data through a web service, with an access link provided in the abstract metadata section. As the data model evolves, OBIS-AU will adapt the workflow to represent data in line with standards.

OBIS-AU engagement with many regional researchers and data providers has demonstrated the potential of centralised data publication. OBIS-AU's success rests in our data transformation workflow and the continual improvement of internal tools. The process involves extensive communication with researchers and data custodians, determining open data licensing, acquiring data, and performing rigorous data quality checks before publication through the Integrated Publishing Toolkit (IPT) (Robertson et al. 2014) to OBIS and GBIF.

Our approach also introduces certain challenges, including researcher concerns regarding accurate data attribution and ownership, a lack of awareness of services offered by OBIS-AU, occasional requests from providers for data embargo, and high demands on limited node resources. To address these challenges, OBIS-AU has intensified its efforts to enhance communication and collaboration with local data providers. This includes participation in conferences and outreach to data managers, researchers, and students. OBIS-AU aims to enrich the Australian biodiversity data publishing landscape by promoting resources such as the OBIS manual, training sessions via the Ocean Teacher Global Academy (OTGA), OBIS GitHub repositories, and offering support to facilitate self-sufficient data standardisation and publication.

OBIS-AU aims to promote the broad adoption of DwC and related data models for publishing Australian marine biodiversity data, fostering a more integrated and accessible global marine biodiversity data environment. Our efforts aim to develop greater autonomy in data publishing within the marine scientific community, advancing the use of standardised data formats crucial for global biodiversity conservation efforts.

Keywords

Darwin Core, WORMS, CSIRO, NCMI, MNF, GBIF

Presenting author

Sachit Rajbhandari

Presented at

SPNHC-TDWG 2024

Acknowledgements

We acknowledge the use of the CSIRO Marine National Facility and the Ocean Biodiversity Information System, Intergovernmental Oceanographic Commission of UNESCO, in undertaking this work.

Hosting institution

Commonwealth Science and Industrial Research Organisation

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Abarenkov K, Andersson A, Bissett A, Finstad AG, et al. (2023) Publishing DNA-derived data through biodiversity data platforms, v1.3. Copenhagen: GBIF Secretariat https://doi.org/10.35035/doc-vf1a-nr22
- Belbin L, Chapman A, Wieczorek J, Morris PJ, Zermoglio PF (2020) Task Group 2 - Data Quality Tests and Assertions. Biodiversity Information Science and Standards 4: 58982. https://doi.org/10.3897/biss.4.58982
- De Pooter D, Appeltans W, Bailly N, et al. (2017) Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5 https://doi.org/10.3897/bdj.5.e10989
- Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, Otegui J, Russell L, Desmet P (2014) The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLoS ONE 9 (8). https://doi.org/10.1371/journal.pone.0102623
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1). https://doi.org/10.1371/journal.pone.0029715